

Geochemical Characterization Using Geophysical Data: a Case Study at the South Oyster Bacterial Transport Site in Virginia

Jinsong Chen^{1*}, Susan Hubbard¹, Yoram Rubin², Chris Murray³, Eric Roden⁴, and Ernest Majer¹

1: Lawrence Berkeley National Laboratory
MS 90-1116, 1 Cyclotron Road, Berkeley, CA 94720

2: Department of Civil and Environmental Engineering
University of California at Berkeley, CA 94720

3: Pacific Northwest National Laboratory
P.O. Box 999, MS K6-81, Richmond, WA 99352

4: Department of Biological Sciences
University of Alabama, Tuscaloosa, AL 354872

Corresponding author: Jinsong Chen

Tel: (510) 486-6842

Fax: (510) 486-5686

Email: jchen@lbl.gov

Submitted to *Environmental Science & Technology*

December, 2002

Abstract

The spatial distribution of field-scale geochemical parameters such as extractable Fe(II) and Fe(III) influences microbial processes and thus the efficacy of bioremediation. Because the traditional characterization of those parameters is invasive and laborious, it is rarely performed sufficiently at the field-scale. As both geochemical and geophysical parameters often correlate to some common physical properties such as lithofacies and clay content, we investigated the utility of geophysical data for improving estimation of geochemical parameters. To our knowledge, this is the first effort to use noninvasive and cost-effective geophysical data to aid in estimation of field-scale geochemical parameters. The data used in the study include physical, geophysical, and geochemical measurements collected in and between several boreholes at the DOE South Oyster Bacterial Transport Site in Virginia. Results show that geophysical data, constrained by physical data, could provide information about extractable Fe(II) and Fe(III) in a minimally invasive manner and with high spatial resolution, which are unparalleled by other geochemical characterization methods.

Introduction

Field-scale bacterial transport in the subsurface plays an important role in the protection of groundwater supplies from contamination by pathogens and in the bioremediation of various metal and radionuclide contaminants [1]. Many studies have shown that both in-situ physical and chemical heterogeneity controls field-scale bacterial movements in saturated aquifers by diverse mechanisms [2,3]. For those bacteria with low or neutral surface charges, the adhesion of bacteria to sediment is primarily determined by physical heterogeneity, specifically by grain size and pore-throat size distributions [4]. However, for those bacterial strains bearing high negative surface charges and traveling through quartz-bearing sediments that contain Fe and Al oxyhydroxides, chemical heterogeneity becomes important due to electrostatic interactions. For example, in sub-neutral pH groundwater, quartz grains have negative charges and metal oxyhydroxides have positive charges, and thus bacterial strains with negative charges will be repelled by quartz grains that lack metal oxyhydroxide coatings but will be attracted to the surfaces of quartz grains that are dominated by metal oxyhydroxides.

In aquifers containing Fe(III) oxides, the activity of dissimilatory iron-reducing bacteria (DIRB) has the potential to influence bacterial transport processes [5]. The adsorption or precipitation of Fe(II) produced by Fe(III) reduction on Fe(III) oxides and on the surface of DIRB may reduce the adhesion of DIRB to sediment grains by reducing the bio-accessible Fe(III) surface areas. Fe(III) oxide reduction could also locally increase the pH of groundwater, and thereby promote the desorption of DIRB by reducing the positive surface charges of the metal oxyhydroxide minerals. In addition, the desorption of daughter cells could increase the net transport rates [6].

Despite the importance of field-scale physical and geochemical heterogeneity on microbial processes, characterization of these parameters remains a daunting task. Traditional methods for

characterizing those parameters, such as laboratory measurements of cores collected from boreholes, are invasive and laborious, and thus are rarely performed sufficiently at the field-scale. In addition, core data are point-scale measurements, which often lead to aliasing of the data. In our previous study, we have successfully estimated the high-resolution spatial distribution of field-scale physical parameters at the DOE South Oyster Bacterial Transport Site in Virginia using geophysical data [7,8]. Comparison of field tracer experiment measurements and numerical modeling predictions, based on geophysically obtained estimates, suggested that the estimated physical parameters provided information at a reasonable scale and resolution for greatly improving the prediction of field-scale solute transport [9].

Motivated by the successful estimation of physical properties using geophysical data, we explore in this study the use of the same noninvasive and cost-effective geophysical data for characterization of geochemical heterogeneity using data collected from the South Oyster site. To our knowledge, this is the first effort to estimate field-scale geochemical parameters using crosshole tomographic geophysical data. We focus on estimation of solid-phase Fe(II) and Fe(III) concentrations along a two-dimensional cross section. The geochemical parameters are known at the borehole locations but are unknown between the boreholes. The two-dimensional spatial distributions of Fe(II) and Fe(III) concentrations are estimated from the crosshole geophysical data using a statistical model. These estimates provide useful information for predicting field-scale microbial Fe(III) oxide reduction potential, and for understanding the field-scale bacterial transport experiments carried out at the South Oyster Focus Area [10].

Site Information and Data

Field Site: The South Oyster site is located on the southern Delmarva Peninsula situated on the eastern coast of the United States between the Chesapeake Bay and Atlantic Ocean. The surficial unconfined aquifer underlying the study area consists of unconsolidated to weakly cemented,

well-sorted, medium- to fine-grained sands and pebbly sands, and the water table is located approximately 2m below the ground surface. A field-scale experiment was undertaken by a multi-disciplinary research team within the uncontaminated aquifer at the site to evaluate the importance of chemical and physical heterogeneity in controlling bacteria that were injected into the ground for bioremediation purposes [3,11]. Two focus areas within the South Oyster Site exist, the Narrow Channel Focus Area and the South Oyster Focus Area (SOFA). Forced gradient chemical and bacterial tracer test experiments were performed within both focus areas.

This study focuses on data collected within a saturated aquifer at a site located ~15m along geological strike from the SOFA transport site, along a transect connecting wells D1, D2, and D3 (Figure 1). Data at this site were collected to investigate the spatial heterogeneity of geochemical parameters and microbial iron reduction potential as an analogue to the immediately adjacent SOFA site. They include the laboratory measurements of physical, geological, geochemical, and biogeochemical parameters from the cores collected at wells D1, D2, and D3 with a sampling interval of 0.15m to 0.30m. The samples were taken between depths of 2.4m and 8.7m below the ground surface [10]. The data also include ground-penetrating radar (GPR) and seismic tomograms along the cross section between wells D1 and D3. In the following, we only describe and analyze the data related to estimation of geochemical parameters, including lithofacies, GPR tomographic attenuation, and solid-phase Fe(II) and Fe(III).

Lithofacies: Borehole lithofacies data were obtained during the logging and core-sampling process based on visual grain size estimation using a comparator chart, and based on soil color and texture. In the original data, four categories were identified: peat, mud, muddy-sand, and sand. According to cluster analysis of physical properties, we reduced the classifications from four to two categories, by grouping peat and mud together and referring to it as mud and by grouping muddy-sand and sand together and referring to it as sand.

The spatial structure of lithofacies was investigated using variograms, which are defined as the average squared difference of a quantity at two locations as a function of the measurement separation distance [12]. We performed variogram analysis by first computing experimental variograms along the vertical direction using data collected from wells D1, D2, and D3, and then fitting it as an exponential model with an integral scale (I_v) of 0.5m. Similarly, we attempted to fit a model to the experimental variogram in the horizontal direction, but with data from just three wells, no model could be reliably fit to the horizontal variogram due to sparse sample density in that direction.

GPR Attenuation: GPR is a geophysical tool that has become increasingly popular as researchers across a variety of disciplines strive to better understand near-surface conditions. GPR uses electromagnetic energy at frequencies of 50-1500 MHz to probe the subsurface. At the frequencies used, the electromagnetic signals propagate primarily as waves. The response of radar is a function of dielectric constant and electrical conductivity of the medium. Radar velocities are most sensitive to the dielectric constant, whereas radar amplitudes are sensitive to both the dielectric constant and the electric conductivity of the medium [13]. Fine-grained soils such as silt and clay have a very strong influence on the electrical conductivity of the medium, and higher radar attenuation (or lower amplitudes) are often associated with high clay content. We hypothesize that both GPR attenuation and geochemical parameters will be influenced by lithology, and focus herein on the use of GPR attenuation to estimate geochemical parameters by exploring the lithology link.

GPR tomographic data were collected along the cross sections between wells D1 and D2 and between wells D2 and D3, using a central frequency of 100MHz. A typical crosshole tomographic geometry consists of two vertical boreholes separated by an inter-well region of interest. Direct energy from a transmitting antenna in one borehole is recorded by the receiving

antenna located in the other borehole. By moving the transmitting and receiving antennas in the boreholes, a large number of ray paths can be recorded that provide a tomographic image of the region between the boreholes [14]. The recorded data include the direct electromagnetic wave travel time from the transmitter to the receiver passing through the crosshole region and the amplitude of the direct arrivals. After dividing the inter-well area into a grid of pixels (0.25m x 0.25m), an inversion algorithm was used to transform the recorded travel time and amplitude information into estimates of the GPR velocity and attenuation at each pixel [15]. Figure 2(a) shows a contoured GPR attenuation tomogram along the cross section between wells D1 and D3. GPR attenuation data ‘overlap’ with borehole data at or near the borehole locations. These collocated data were used to develop site-specific relations between geophysical attributes and geochemical parameters, which enabled us to extend geochemical measurements from the boreholes to the inter-well area using the tomographic data.

At our study site, statistical analysis revealed that GPR attenuation has a good correlation with lithofacies. The GPR amplitudes were more attenuated when passing through mud than through sand. Based on data collected at the three wells, the mean GPR attenuation of mud is 0.75 1/m with a standard deviation of 0.30 1/m, whereas the mean GPR attenuation of sand is 0.49 1/m with a standard deviation of 0.11 1/m. This correlation between GPR attenuation and lithofacies is reasonable. GPR attenuation under low-loss conditions is determined by both sediment dielectric constant and electrical conductivity. In unsaturated aquifers, dielectric constant is a major factor due to its sensitivity to changes in water content. However, in saturated aquifers such as the one under considered here, GPR attenuation is usually dominated by electrical conductivity. Since mud includes a large portion of silt and clay, which have much higher electrical conductivity than sand, GPR attenuation in saturated aquifers is expected to have larger values in mud than in sand [13].

Extractable Fe(II): Extractable Fe(II) was measured by leaching the triplicate 0.5-1.0g subsamples, obtained from each depth interval of wells D1, D2, and D3, with 0.5M HCl for one hour [16]. The extracted Fe(II) may include both structural Fe(II) from native iron-bearing minerals in the formation and reduced Fe(II) from bacterial Fe(III) reduction. Figure 3(a) shows the histogram of the logarithmic extractable Fe(II) based on data collected at wells D1, D2, and D3. It appears to be symmetrically distributed around the mean.

Statistical analysis revealed that Fe(II) also has a good correlation with lithofacies, and that mud has much higher concentrations of Fe(II) compared to those of sand. Based on Fe(II) data collected at the three wells, the mean logarithmic Fe(II) of mud is 3.36 ($\mu\text{mol/cc}$) with a standard deviation of 1.19 ($\mu\text{mol/cc}$), whereas the mean logarithmic Fe(II) of sand is only 0.40 ($\mu\text{mol/cc}$) with a standard deviation of 1.72 ($\mu\text{mol/cc}$). The Fe(II) distribution as a function of lithofacies is reasonable as Fe(II) is usually sequestered in fine-grain sediments such as silt and clay [1], and our mud lithofacies includes a large portion of silt and clay .

As extractable Fe(II) and GPR attenuation were both correlated with lithofacies, we have physical justification for investigating the link between Fe(II) and GPR attenuation. However, the connection is complex and may be affected by organic matter, grain size, porosity, and other physical parameters. Figure 4 shows the cross-plot of logarithmic Fe(II) versus logarithmic GPR attenuation based on data at wells D1, D2, and D3. In the figure, the circles represent sand and the triangles represent mud. For sand, extractable Fe(II) reveals a strong positive correlation with GPR attenuation, and logarithmic Fe(II) linearly increases with increasing logarithmic GPR attenuation. For mud, however, this correlation is less reliable, and the increment of logarithmic Fe(II) is not apparent as logarithmic GPR attenuation increases. This may be because mud has a high clay content that includes rich organic matter, which makes the relationship between GPR attenuation and clay content less predictable. Based on the results, we developed a statistical

model, as will be described in the next section, for extractable Fe(II) estimation using only GPR attenuation and lithofacies information. In this model, logarithmic Fe(II) was considered as a linear function of logarithmic GPR attenuation for sand but as a constant for mud.

Extractable Fe(III): Extractable Fe(III), including amorphous and crystalline Fe(III) oxides, was determined using a citrate-dithionite reagent (pH 4.8) [17]. Figure 3(b) shows the histogram of logarithmic extractable Fe(III) based on data collected at wells D1, D2, and D3. Unlike logarithmic extractable Fe(II), the logarithmic extractable Fe(III) is skewed toward the larger values.

Data analysis suggested that Fe(III) has weak correlations with all of the physical properties. To further identify main factors that affect extractable Fe(III), we considered a full linear model with factors including depth, Fe(II), lithofacies, clay content, silt fraction, mean grain size, grain sorting, porosity, and organic matter. The full model should give the maximum squared multiple correlation coefficient, which measures the goodness of fit of the model. When depth was excluded from the full model, the multiple squared correlation coefficient is dramatically reduced, from 0.67 to 0.36. With the drop of Fe(II) from the full model, the coefficient is also reduced significantly, from 0.67 to 0.55. For all other factors, however, the multiple squared correlation coefficients was only reduced slightly or not at all. This analysis suggested that Fe(II) and depth were the most significant exploratory variables for Fe(III).

The correlation between Fe(III) and Fe(II) under iron reducing conditions is reasonable, considering microbial reduction mechanisms [18]. Fe(III) serves as an electron acceptor, which is required for iron reduced bacterial cell growth and maintenance, while Fe(II) is the direct product of the iron reduction [1]. In the laboratory and under strictly controlled conditions, Fe(III) and Fe(II) can be connected through deterministic empirical models [19]. However, the concentration of extractable Fe(III) is also correlated to depth within the sedimentary sequence. The dependence on depth may be related to changes in lithology, because the grain size of

sediment in sand is finer in the deeper part of the section. The finer grain size in the deeper sediment would be associated with greater surface areas, which may explain the greater concentration of extractable Fe(III). There may also be a net downward flow at the site, which could transport Fe(II) in solution from the mud/peat layer in the middle of the section. The Fe(II) could then be oxidized in the lower sandy sequence below the mud/peat, where the concentration of dissolved oxygen is greater. However, with the available data, we are unable to assess the extent to which the solid-phase Fe(III) and Fe(II) distributions are controlled by ongoing microbial reduction processes in the deposit, or by slow redistribution of iron during coupled transport/reaction processes acting on materials whose original composition was determined by short-term events taking place at the time of deposition. This uncertainty does not negate the usefulness of the analysis presented here, as our goal is to advance methodologies for linking sediment geochemical properties to geophysical data, rather than to provide a mechanistic explanation for the origin of the geochemical properties.

Statistical Estimation Methods

In this section, we develop a statistical model to estimate extractable Fe(II) and Fe(III) along a two-dimensional cross section that traverses wells shown in Figure 1, using the spatial correlation of lithofacies and the cross correlations among Fe(II), Fe(III), lithofacies, and GPR attenuation, obtained in the proceeding section. We first divided the cross section (12m x 6m) into 1225 pixels, which have the dimensions of 0.25m x 0.25m. Lithofacies, Fe(II), and Fe(III) were known at pixels along wells D1, D2, and D3 but were unknown at other pixels located between the wells. GPR attenuation was considered ‘known’ at all locations. Our goal in this study is to estimate extractable Fe(II) and Fe(III) at each pixel using data available at the boreholes and along the cross section.

Statistical model: Figure 5 shows the structure of the developed statistical model. Lithofacies, logarithmic Fe(II) concentrations, and logarithmic Fe(III) concentrations at pixel- i were considered as random variables and represented as $l(i)$, $f2(i)$, and $f3(i)$, respectively. GPR attenuation and depth at each pixel were considered as data and were represented as $a(i)$ and $d(i)$. Lithofacies $l(i)$ at pixel- i is an indicator random variable taking on two possible values: 1 for sand and 0 for mud. The probability of sand being present at a particular pixel was obtained by conditioning the estimate to the lithofacies of the adjacent pixels using the lithofacies spatial correlation model, and to the collocated GPR attenuation and logarithmic Fe(II) concentrations using their cross correlations. GPR attenuation at pixel- i was assumed to have a normal distribution with a mean $[\alpha_1 \cdot l(i) + \alpha_2 \cdot (1 - l(i))]$ and a standard deviation σ_1 . The standard deviation is a measure of errors associated with GPR attenuation data acquisition and inversion. If the lithofacies at pixel- i is sand, i.e. $l(i) = 1$, the conditional mean GPR attenuation is α_1 , and otherwise it is α_2 .

Both extractable Fe(II) and Fe(III) were considered as dependent random variables. A logarithmic Fe(II) concentration at each pixel was conditioned to the collocated lithofacies and GPR attenuation value. The conditional probability of Fe(II) given lithofacies and GPR attenuation was assumed to have a normal distribution with a mean value $[(\beta_1 + \beta_2 \cdot a(i)) \cdot l(i) + \beta_3 \cdot (1 - l(i))]$ and a standard deviation σ_2 . The standard deviation accounts for uncertainty associated with the cross correlation between GPR attenuation and Fe(II). Based on the previous data analysis (Figure 4), the mean of logarithmic Fe(II) concentrations in the above expression was modeled as a linear function of GPR attenuation for sand but as a constant for mud. The logarithmic Fe(III) concentration at each pixel was modeled to depend on the collocated Fe(II) concentration and on depth, but considered to be independent of lithofacies and GPR attenuation, given the collocated Fe(II) concentration. The conditional

probability of Fe(III) was assumed to have a normal distribution with a mean $[\gamma_1 + \gamma_2 \cdot f2(i) + \gamma_3 \cdot d(i)]$ and a standard deviation σ_3 .

Parameters α_1 , α_2 , β_1 , β_2 , β_3 , γ_1 , γ_2 , and γ_3 are coefficients of cross correlations among lithofacies, GPR attenuation, Fe(II), and Fe(III). These parameters were obtained reliably from regression analysis of data collected at wells D1, D2, and D3. Although we can also obtain the estimates of parameters σ_1 , σ_2 , and σ_3 from the same analysis, they are less reliable compared to the estimates of those coefficients. As a result, we considered α_1 , α_2 , β_1 , β_2 , β_3 , γ_1 , γ_2 , and γ_3 as known parameters but σ_1 , σ_2 , and σ_3 as unknown parameters in our model.

Parameters I_v and I_h represent the integral lengths of lithofacies along the vertical and horizontal directions, respectively. In this investigation, both of them were considered as known parameters within our model in order to simplify model structure. As previously discussed, we can virtually obtain a good estimate of the vertical integral length ($I_v=0.5\text{m}$) by performing variogram analysis. However, the data were not dense enough for reliably inferring the horizontal integral length. Instead, we used 3.0m as an initial guess of I_h and later varied the value over a range from 0.5m to 6.0m with an increment of 0.5m to see how sensitive the estimated Fe(II) and Fe(III) values were to the choice of the horizontal integral length I_h . The sensitivity analysis showed that the estimated Fe(II) and Fe(III) were not sensitive to the value of I_h as long as it was greater than 0.5m. This is possibly because part of spatial correlation of lithofacies was implicitly conveyed by spatially exhaustive GPR attenuation.

Parameters σ_1 , σ_2 , and σ_3 were considered as random quantities and assigned prior distributions using a Bayesian prospective [20]. In light of the fact that parameters σ_1 , σ_2 , and σ_3 are nonnegative and to avoid improper posterior distributions, we instead used the inverse

squares of those parameters and assumed their prior distributions were of the form of gamma distributions. Using the previously assigned conditional probabilities of lithofacies, logarithmic Fe(II), logarithmic Fe(III), and the assumed prior distributions of parameters σ_1 , σ_2 , and σ_3 , we derived conditional distributions for all the random variables and for the unknown parameters σ_1 , σ_2 , and σ_3 .

Monte Carlo Sampling: As extractable Fe(II) and Fe(III) at each pixel are random variables, our goal is to estimate their distributions, from which we can calculate their mean, variance, and other statistics. Fe(III) at each pixel correlates to Fe(II), and Fe(II) depends on GPR attenuation and lithofacies, which spatially correlates to lithofacies at other pixels. Consequently, all the random variables and the unknown parameters are linked. The total number of random quantities, including all the random variables and the unknown parameters, is about 3500. Given such a large number, it is prohibitive to estimate the probability distributions of those variables using analytical methods. Instead, we used a Markov chain Monte Carlo method to generate hundreds of realizations for each variable, from which we can calculate mean, variance, and even its probability density function.

Different from the traditional Monte Carlo methods, which generate independent samples from conditional distributions, the Markov chain Monte Carlo methods generate dependent samples by running a cleverly constructed Markov chain from given starting values. As the initially generated realizations depend on the starting values, we will not use them for inference in order to avoid the bias in choosing the starting values. The remaining generated realizations were stationary and thus used to calculate mean, variance, quantiles, and other statistics, which should converge to their corresponding asymptotic values with the increase of the number of realizations, as dictated by the ergodic theorem [20]. The Markov chain Monte Carlo methods

provided us with a practical tool to solve our estimation problem, which involves a large number of random variables.

As many variables in our model are conditionally independent, we used the Gibbs sampling method to generate Markov chains. The sampling method involved the following sequences: (1) specifying probability distributions for each variable and data, (2) deriving conditional distribution for each variable, and (3) drawing samples from the conditional distributions. The sampling process is sequential. For example, when we draw a sample from the conditional probability of σ_1 , we assumed all other variables and parameters are given. Next, when we generate a sample from the conditional probability of σ_2 , we consider the previously generated σ_1 as the true value. The above process was repeated for many times so that we obtained 2400 samples for each variable. We decided to reject first 400 samples for each variable, as these values are more likely depend on the choice of the starting values. We then used the remaining 2000 samples to compute mean, variance, and quantiles of the variable.

Cross validation: The effectiveness of the developed model was evaluated through cross validation using data at wells D1, D2, and D3. Each of those wells is in turn considered as a testing well and the other wells as training wells. We first used data at the training wells and GPR attenuation data along the transect from wells D1 to D3 to estimate Fe(II) and Fe(III) concentrations at the testing locations, and then compared the estimated results with their corresponding true values to evaluate the model. For example, using data from wells D1 and D3 only and the GPR data, we estimated extractable Fe(II) and Fe(III) at all locations and validated the estimates at well D2. This procedure was repeated for all other possible well-pair configurations.

Results and Discussion

Fe(II) estimation: Using the above statistical model and the Gibbs sampling method, we estimated extractable Fe(II) at each well using data from the other two wells. Figure 6(a) compares the estimated mean of logarithmic Fe(II) concentrations (solid black lines) with their corresponding core measurements (red circles with red solid lines) at testing wells D1, D2, and D3, respectively. The dashed lines indicate 95% confidence intervals, which means if we take many repeated measurements at the same location, 95% of those measurements should fall within the range. The figures suggest that the developed model is effective for logarithmic Fe(II) estimation. As shown in the figures, the mean estimates of logarithmic Fe(II) at well D1 closely follow the true measurements of logarithmic Fe(II), and the mean estimates of logarithmic Fe(II) at testing well D2 has a good agreement with the true values. Although the estimated results of logarithmic Fe(II) at testing well D3 is not as good as at other testing wells, most measurements are still within the 95% confidence intervals of the estimated values. To show the improvement offered by the GPR data, Figure 6(b) shows the estimated mean of logarithmic Fe(II) without using GPR attenuation. Although the three wells in the cross section are separated by only about 6m, the estimated results without using GPR attenuation are much worse than those obtained with the use of GPR attenuation. The mean estimates of Fe(II) are smooth and the 95% confidence intervals are much larger than the model including GPR attenuation.

Using lithofacies and Fe(II) data at wells D1, D2, and D3 and the GPR attenuation data along the cross sections from wells D1 to D3, we estimated extractable Fe(II) along the cross section between wells D1 and D3. Figure 2(b) shows the estimated mean of logarithmic Fe(II) along the two-dimensional transect. The figure provides detailed spatial information about extractable Fe(II) on the cross section, which cannot be obtained from borehole measurements only. The

estimated spatial pattern of Fe(II) is similar to that of GPR attenuation because of the good correspondence of Fe(II) with GPR attenuation.

As the byproduct of the Fe(II) estimation, we obtained the probability of sand occurring at each pixel on the cross section between wells D1 and D3 (Figure 2(c)). Lithofacies on the cross section between wells D2 and D3 corresponds to GPR attenuation very well, with sand having lower attenuation and mud having higher attenuation. However, lithofacies on the lower part of the cross section between wells D1 and D2 does not correspond directly to GPR attenuation, where GPR attenuation is high but the estimated lithofacies is sand. There are two possible reasons. One is that lithofacies at each pixel was determined by both lithofacies measurements at boreholes and GPR attenuation along the cross section. The other is that our sand lithofacies classification includes both pure sand and muddy-sand components, and lithofacies at those pixels in reality may be muddy-sand rather than pure sand, which would tend to have a higher GPR attenuation than pure sand.

Fe(III) estimation: Following a similar process as described above, we estimated Fe(III) and performed a cross validation analysis. Figure 7 shows the estimated mean of logarithmic Fe(III) and their 95% confidence intervals at each testing well. Again, Fe(III) estimates at each testing well were obtained using data at the other two training wells, along with the GPR attenuation. We found that the measured logarithmic Fe(III) values followed the trends of the estimated mean, and that most of the estimates were within the 95% confidence intervals.

Using the calibrated Fe(III) model, we also estimated logarithmic Fe(III) at each pixel along the cross section between wells D1 and D3 (Figure 2(d)). The two-dimensional image of Fe(III) is much different from the GPR attenuation image. Comparison of Figures 2(c) and 2(d) reveals that extractable Fe(III) has higher concentrations beneath the mud/peat layer than above the layer.

Discussion: The developed approach is important, given that currently no other methods can sufficiently predict the spatial distribution of field-scale extractable Fe(II) and Fe(III) at both a reasonable cost and in a minimally invasive manner. Although the model cannot accurately estimate Fe(II) and Fe(III) point values at the core-scale, it provided estimates at the spatial resolution of the tomographic data (0.25m x 0.25m for this study), which was previously determined to be an effective characterization scale for understanding field-scale chemical and bacteria transport [3,9]. The estimated spatial distributions of Fe(II) and Fe(III) may be used for the inference of microbial iron reduction potential [21] and for the estimation of bacterial attachment or detachment parameters, which are needed as input for numerical predictions of subsurface bacterial transport.

The developed model is based on the data analysis results without considering field-scale geochemical and biogeochemical processes, and thus is site-specific. However, our hypotheses that geochemical parameters and geophysical attributes can be linked through their mutual dependence on physical properties may be applicable to estimation of other geochemical parameters, such as organic matter and other metal oxyhydroxides, as well as to other field sites.

Figure Captions

Figure 1: Locations of the South Oyster Bacterial Transport Site and our study area (D-site).

Figure 2: (a) GPR attenuation; (b) Estimated mean logarithmic Fe(II); (c) Frequency of sand, where a frequency of 0.0 implies that lithofacies is mud/peat while a frequency of 1.0 implies that lithofacies is sand/muddy-sand; (d) Estimated mean logarithmic Fe(III).

Figure 3: Histograms of logarithmic Fe(II) and Fe(III).

Figure 4: Cross-plot of logarithmic Fe(II) versus logarithmic GPR attenuation, where the circles represent sand/muddy-sand and the solid triangles represent mud/peat lithofacies. These data suggest that there is a strong correlation between logarithmic Fe(II) and logarithmic GPR attenuation for sand/muddy-sand, but a weak correlation for mud/peat.

Figure 5: Statistical model for Fe(II) and Fe(III) estimation, where $l(i)$, $f2(i)$, $f3(i)$, $a(i)$, and $d(i)$ represent lithofacies, Fe(II), Fe(III), GPR attenuation, and depth at pixel- i , respectively. Parameters α_1 , α_2 , β_1 , β_2 , β_3 , γ_1 , γ_2 , and γ_3 are coefficients of cross correlations among lithofacies, Fe(II), Fe(III), and GPR attenuation. Variables σ_1 , σ_2 , and σ_3 are measures of uncertainty associated with the cross correlations. I_v and I_h are the integral lengths of lithofacies along the vertical and horizontal directions. The rectangles indicate that the parameters are given, while the circles indicate that the variables are unknown.

Figure 6: Comparisons of the true logarithmic Fe(II) and its corresponding estimated values obtained (a) using GPR attenuation, (b) without using GPR attenuation.

Figure 7: Comparisons of the estimated logarithmic Fe(III) and its corresponding true values (a) at testing well D1, (b) at testing well D2, and (c) at testing well D3.

Acknowledgement

This study was supported by the Assistant Secretary for Energy Research, Office of Health and Environmental Research, under the direction of element manager Frank Wobber, and with U.S. Department of Energy contract DE-AC03-76SF00098 to Ernest L. Majer and NSF grant EAR 9628306 to Yoram Rubin. All computations were carried out at the Center for Computational Seismology (CCS), supported by DOE's Office of Basic Energy Science, at Lawrence Berkeley National Laboratory.

References:

1. Chappelle, F. H. *Groundwater Microbiology and Geochemistry*, 2nd Edition, John Wiley & Sons, New York, **2001**.
2. Chen, J.; Hubbard, S.; Rubin, Y.; Murray, C.; Roden, E.; Majer, E. *Eos Transaction*, **2002**, 83.
3. DeFlaun, M.; ...; Chen, J.; Others *EOS Transactions*, **2001**, 82, 417.
4. Dong, H. L.; Onstott, T. C.; DeFlaun, M.; Fuller, M. E.; Scheibe, T.; Streger, S.; Rothmel, R.; Mailloux, B. J. *Environ. Sci. Technol.*, **2002**, 36, 891.
5. Caccavo, F.; Schamberger, P. C.; Keiding, K.; Nielsen, P. H. *Appl. Environ. Microbiol.*, **1997**, 63, 3837.
6. Roden, E. E.; Urrutia, M. M.; Mann, C. J. *Appl. Environ. Microbiol.*, **2000**, 66, 1062.
7. Chen, J.; Hubbard, S.; Rubin, Y. *Water Resources Research*, **2001**, 37, 1603.
8. Hubbard, S.; Chen, J.; Peterson, J.; Majer, E.; Williams, K.; Swift, D.; Mailloux, B.; Rubin, Y. *Water Resources Research*, **2001**, 37, 2431.
9. Scheibe, T. D.; Chien, Y. J. *Ground Water*, **2003**, in press.
10. Murray, C. J.; Xie, Y.; Roden, E. E.; Overstreet, K. *Geological Society of American Annual Meeting, Boston, MA*, **2001**.
11. Johnson, W. P.; Zhang, P.; Fuller, M. E.; Scheibe, T. D.; Mailloux, B. J.; Onstott, T. C.; DeFlaun, M. F.; Hubbard, S. S.; Radtke, J.; Kovacic, W. P.; Holben, W. *Environ. Sci. Technol.* **2001**, 35, 182.
12. Rubin, Y. *Applied Stochastic Hydrology*, Oxford University Press, **2003**, in press.
13. Davis, J. L.; Annan, A. P. *Geophysical Prospecting*, **1989**, 37, 531.
14. Peterson, J. E.; Paulsson, B. N.; McEvelly, T. V. *Geophysics*, **1985**, 50, 1566.
15. Peterson, J. E. *J. Environ. Eng. Geophys.*, **2001**, 6, 1.

16. Lovley, D.R.; Phillips, E. J. P. *Appl. Environ. Microbiol.*, **1986**, *51*, 683.
17. Canfield, D.E. *Geochim. Cosmochim. Acta*, **1989**, *53*, 619.
18. Lovley, D.R. *Microbiol. Rev.*, **1991**, *55*, 259.
19. Liu, C.; Kota, S.; Zachara, J. M.; Fredrickson, J. K.; Brinkman, C. *Environ. Sci. Technol.*, **2001**, *35*, 2482.
20. Gilks, W. R.; Richardson, S.; Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice*, **1996**, Chapman & Hall/CRC.
21. Roden, E. E.; Urrutia, M. M. *Geomicrobiol. J.*, **2002**, *19*, 209.

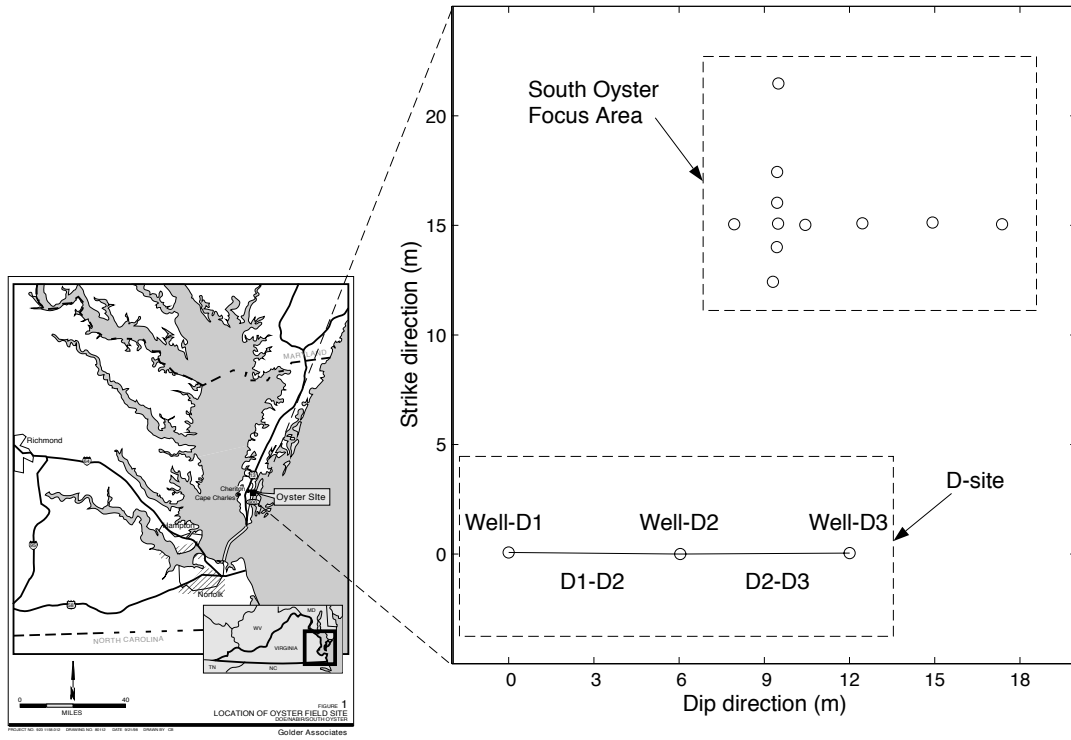


Figure 1: Locations of the South Oyster Bacterial Transport Site and our study area (D-site).

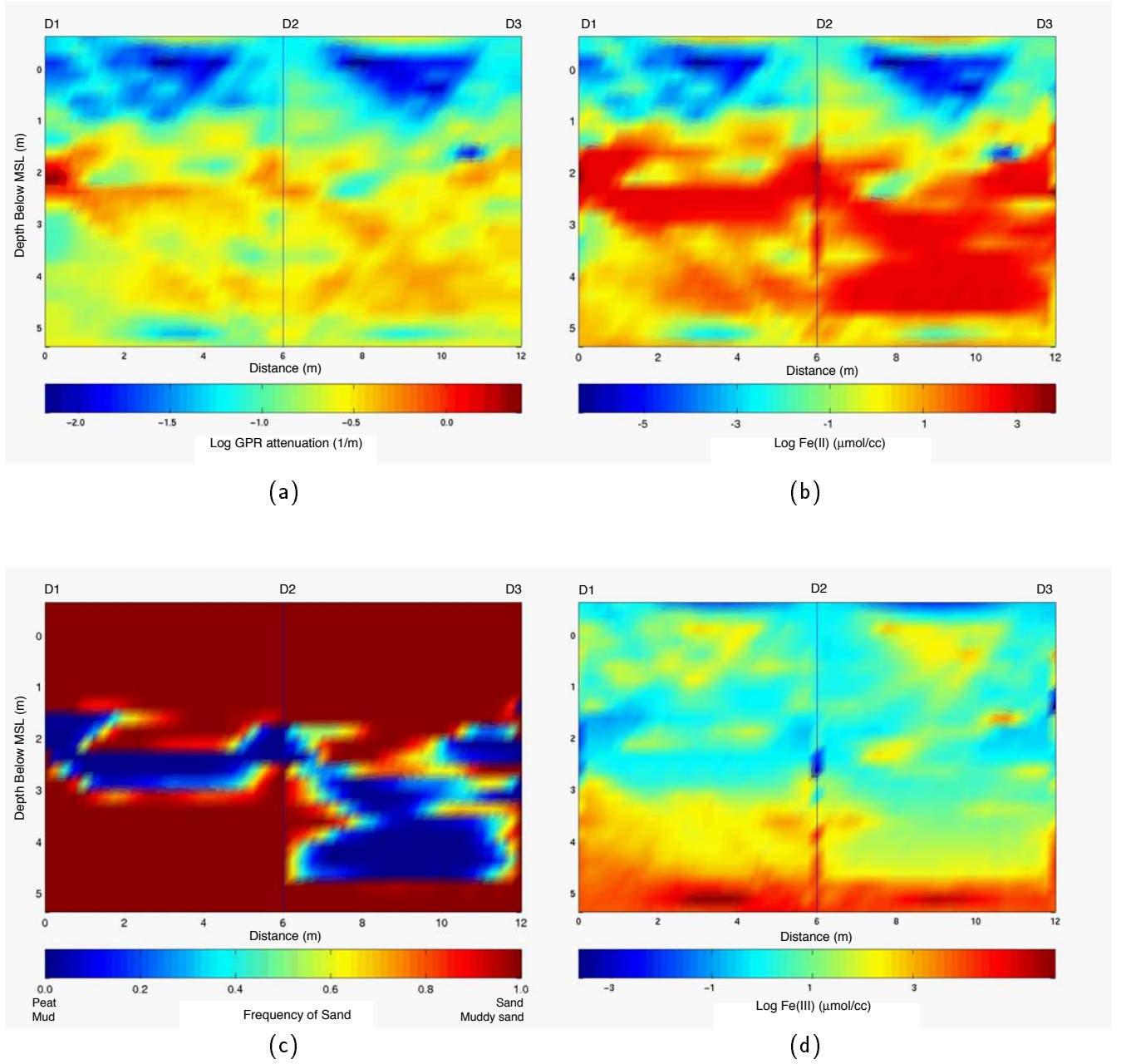


Figure 2: (a) GPR attenuation; (b) Estimated mean logarithmic Fe(II); (c) Frequency of sand, where a frequency of 0.0 implies that lithofacies is mud/peat while a frequency of 1.0 implies that lithofacies is sand/muddy-sand; (d) Estimated mean logarithmic Fe(III).

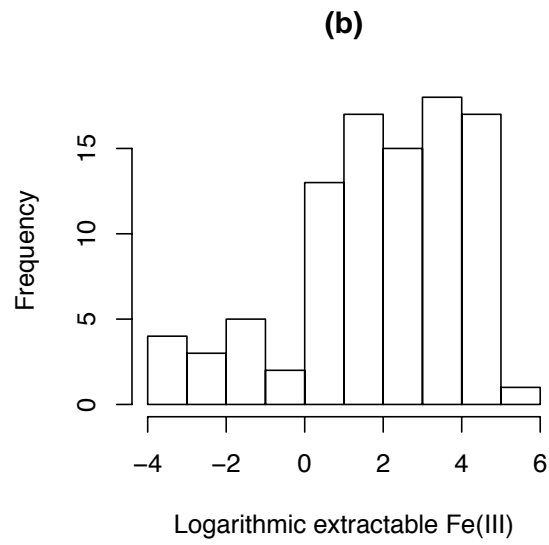
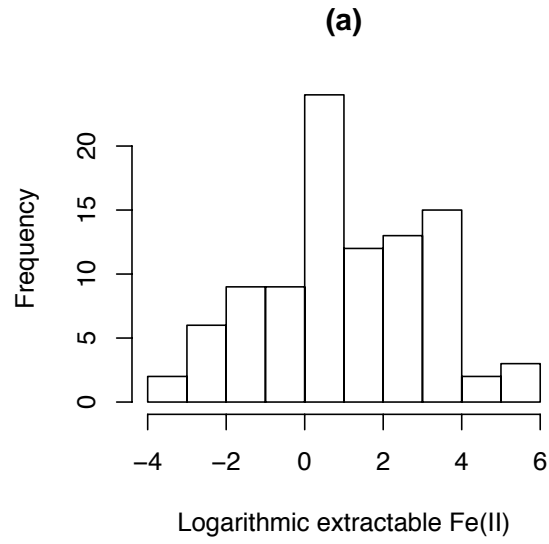


Figure 3: Histograms of logarithmic Fe(II) and Fe(III).

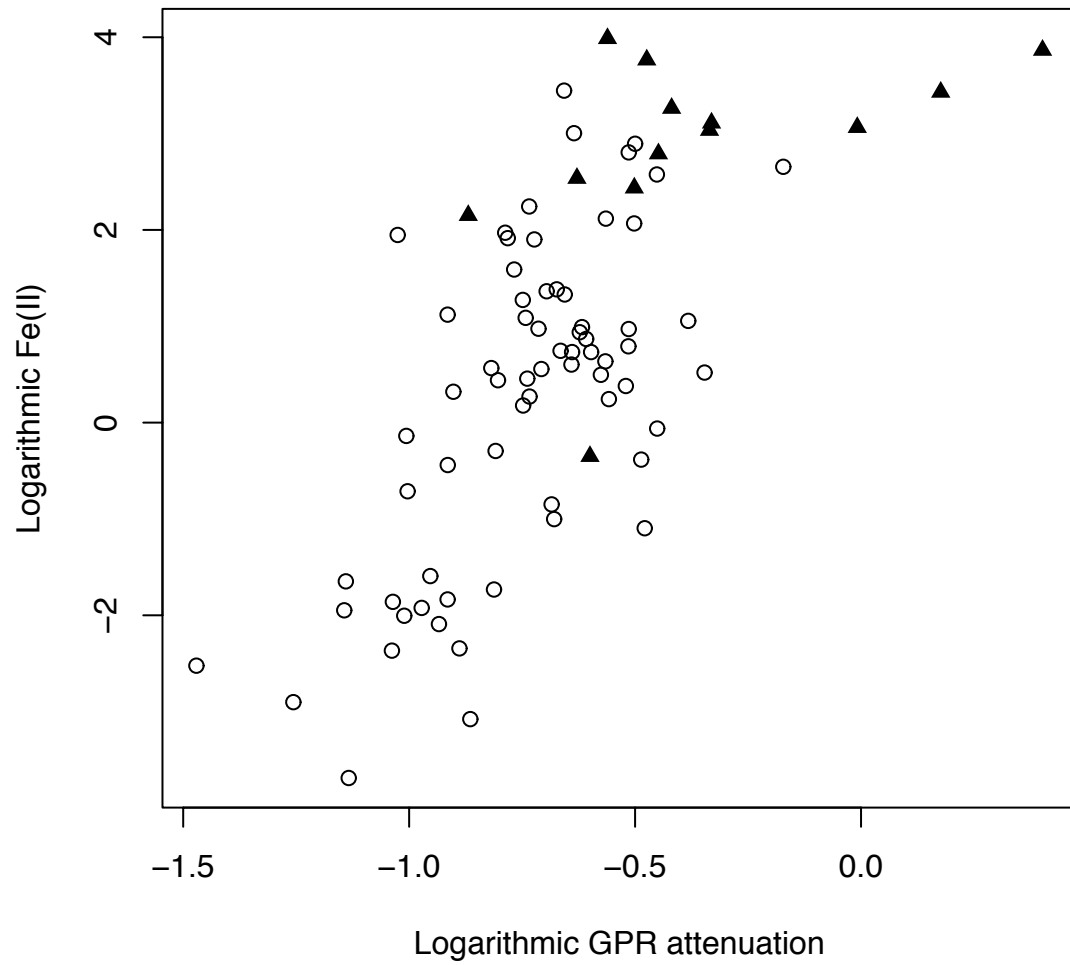


Figure 4: Cross-plot of logarithmic Fe(II) versus logarithmic GPR attenuation, where the circles represent sand/muddy-sand and the solid triangles represent mud/peat lithofacies. These data suggest that there is a strong correlation between logarithmic Fe(II) and logarithmic GPR attenuation for sand/muddy-sand, but a weak correlation for mud/peat.

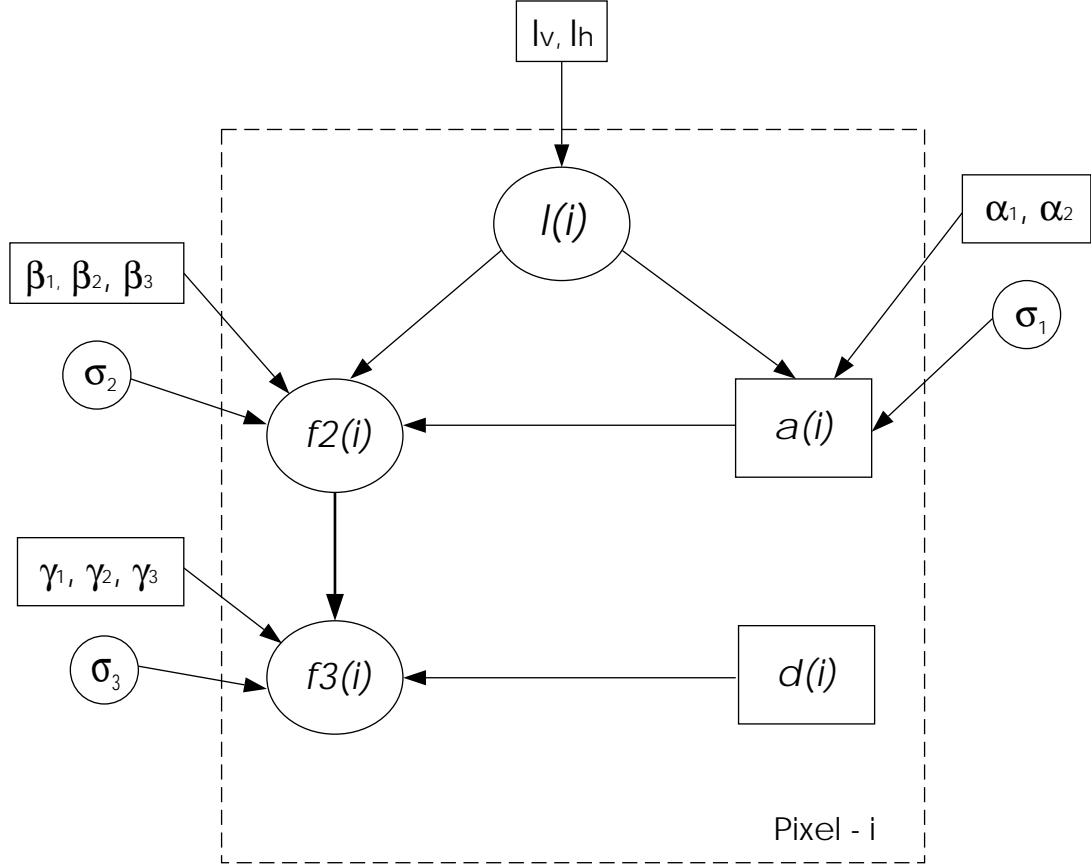


Figure 5: Statistical model for Fe(II) and Fe(III) estimation, where $l(i)$, $f2(i)$, $f3(i)$, $a(i)$, and $d(i)$ represent lithofacies, Fe(II), Fe(III), GPR attenuation, and depth at pixel- i , respectively. Parameters α_1 , α_2 , β_1 , β_2 , β_3 , γ_1 , γ_2 , and γ_3 are coefficients of cross correlations among lithofacies, Fe(II), Fe(III), and GPR attenuation. Variables σ_1 , σ_2 , and σ_3 are measures of uncertainty associated with the cross correlations. I_v and I_h are the integral lengths of lithofacies along the vertical and horizontal directions. The rectangles indicate that the parameters are given, while the circles indicate that the variables are unknown.

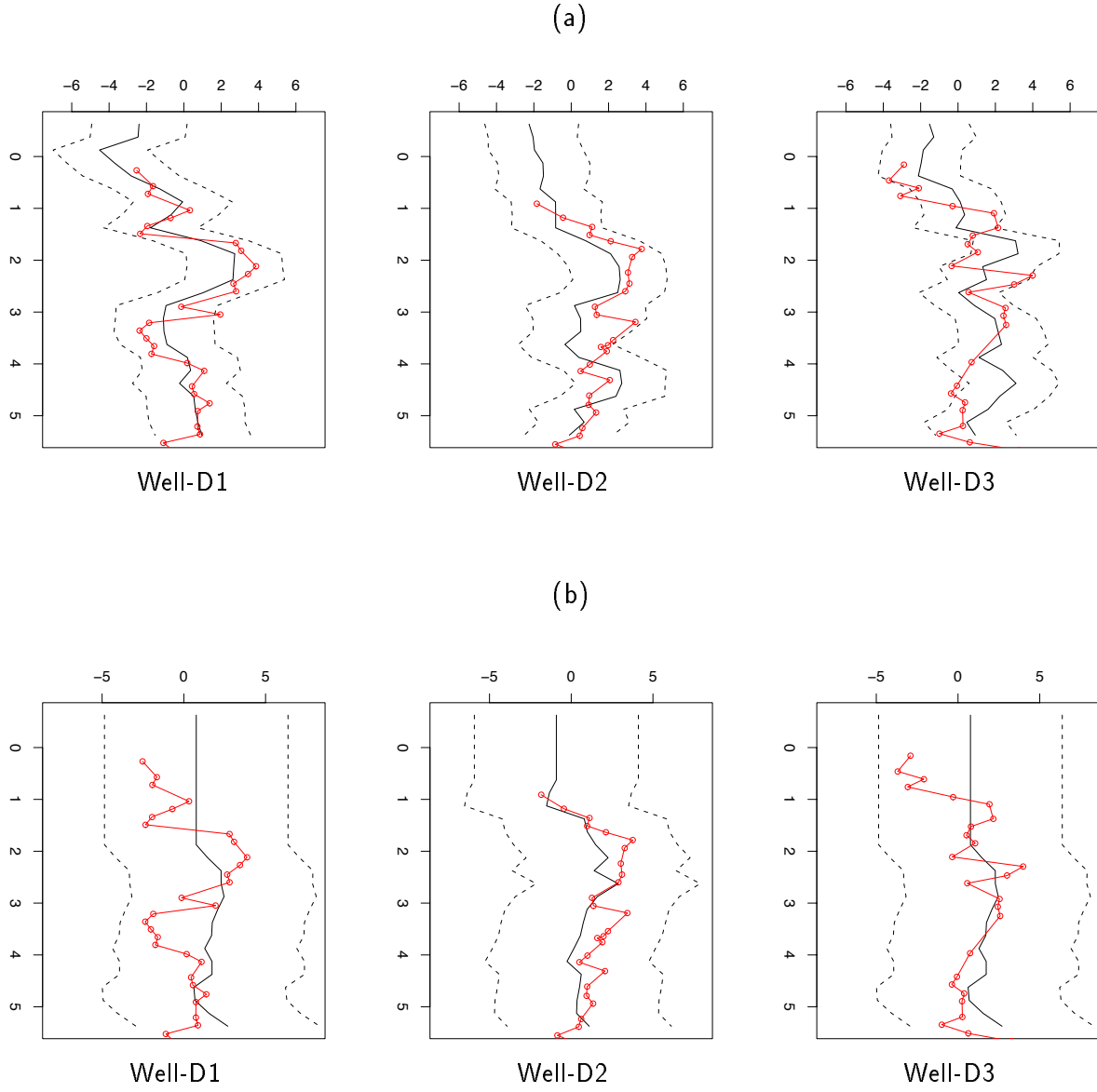


Figure 6: Comparisons of the true logarithmic Fe(II) and its corresponding estimated values obtained (a) using GPR attenuation, (b) without using GPR attenuation.

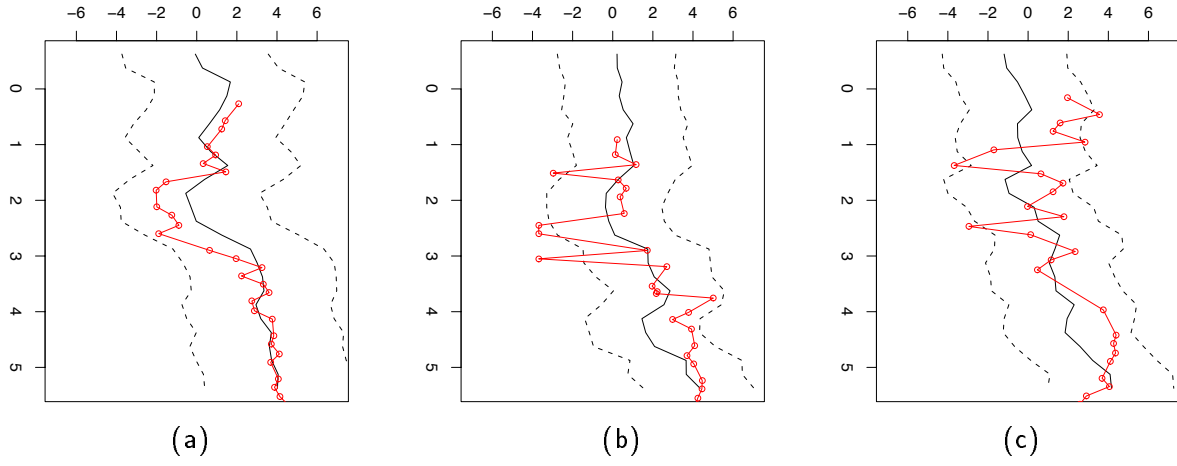


Figure 7: Comparisons of the estimated logarithmic Fe(III) and its corresponding true values (a) at testing well D1, (b) at testing well D2, and (c) at testing well D3.